# Finite-time last-iterate convergence for multi-agent learning in games

Tianyi Lin, Zhengyuan Zhou, Panayotis Mertikopoulos, Michael I Jordan

# Finite-Time Last-Iterate Convergence for Multi-Agent Learning in Games

**Tianyi Lin** [*1] **Zhengyuan Zhou** [*2] **Panayotis Mertikopoulos** [3] **Michael. I. Jordan** [4]

## Abstract

In this paper, we consider multi-agent learning via online gradient descent in a class of games called $\lambda$-cocoercive games, a fairly broad class of games that admits many Nash equilibria and that properly includes unconstrained strongly monotone games. We characterize the finite-time last-iterate convergence rate for joint OGD learning on $\lambda$-cocoercive games; further, building on this result, we develop a fully adaptive OGD learning algorithm that does not require any knowledge of problem parameter (e.g. cocoercive constant $\lambda$) and show, via a novel double-stopping time technique, that this adaptive algorithm achieves same finite-time last-iterate convergence rate as non-adaptive counterpart. Subsequently, we extend OGD learning to the noisy gradient feedback case and establish last-iterate convergence results–first qualitative almost sure convergence, then quantitative finite-time convergence rates– all under non-decreasing step-sizes. To our knowledge, we provide the first set of results that fill in several gaps of the existing multi-agent online learning literature, where three aspects–finite-time convergence rates, non-decreasing step-sizes, and fully adaptive algorithms have been unexplored before.

## 1. Introduction

In its most basic incarnation, online learning (Blum, 1998; Shalev-Shwartz et al., 2012; Hazan, 2016) can be described as a feedback loop of the following form:

1. The agent interfaces with the environment by choosing an *action* $a_t \in \mathcal{A} \subseteq \mathbf{R}^d$ (e.g., bidding in an auction, selecting a route in a traffic network).

2. The environment then yields a reward function $r_t(\cdot)$, and the agent obtains the reward $r_t(a_t)$ and receives some *feedback* (e.g., reward function $r_t(\cdot)$, gradient $\nabla r_t(a_t)$, or reward $r_t(a_t)$), and the process repeats.

As the reward functions $r_t(\cdot)$ are allowed to change from round to round, the standard metric that quantifies the performance of an online learning algorithm is that of regret (Blum & Mansour, 2007): at time $T$, the regret is the difference between $\max_{a\in A} \sum_{t=1}^{T} u_t(a)$, the total rewards achieved by the best fixed action in hindsight, and $\sum_{t=1}^{T} u_t(a_t)$, the total rewards achieved by the algorithm. In the rich online learning literature (Zinkevich, 2003; Kalai & Vempala, 2005; Shalev-Shwartz & Singer, 2007; Arora et al., 2012; Shalev-Shwartz et al., 2012; Hazan, 2016), perhaps the simplest algorithm that achieves the minimax-optimal regret guarantee is Zinkevich's online gradient descent (OGD), where the agent simply takes a gradient step (at current action) to form the next action, performing a projection if necessary. Due to its simplicity and strong performance, it is arguably one of the most widely-used algorithms in online learning theory and applications (Zinkevich, 2003; Hazan et al., 2007; Quanrud & Khashabi, 2015).

At the same time, the most common instantiation of the above online learning model (where reward functions change arbitrarily over time) is multi-agent online learning: each agent is making online decisions in an environment that consists of other agents who are simultaneously making online decisions and whose actions impact the rewards of other agents; that is, each agent's reward is determined by an (unknown) game. Note that in multi-agent online learning, as other agents' actions change, each agent's reward function, when viewed solely as a function of its own action, also changes, despite the fact that the underlying game mechanism is fixed. Consequently, in this setting, the universality of the OGD regret bounds raises high expectations in terms of performance guarantees, leading to the following fundamental question in game-theoretical learning (Cesa-Bianchi & Lugosi, 2006; Shoham & Leyton-Brown, 2008; Viossat & Zapechelnyuk, 2013; Bloembergen et al., 2015; Monnot & Piliouras, 2017): *Would OGD learning, and more broadly no-regret learning, lead to Nash equilibria?*

[*]Equal contribution [1]Department of Industrial Engineering and Operations Research, UC Berkeley [2]Stern School of Business, New York University and IBM Research [3]Univ. Grenoble Alpes, CNRS, Inria, LIG, 38000 Grenoble and Criteo AI lab [4]Department of Statistics and Electrical Engineering and Computer Science, UC Berkeley. Correspondence to: Tianyi Lin <darren_lin@berkeley.edu>.

*Proceedings of the $37^{th}$ International Conference on Machine Learning*, Vienna, Austria, PMLR 119, 2020. 

As an example, if all users of a computer network individually follow some no-regret learning algorithm (e.g. OGD) to learn the best route for their traffic demands, would the system eventually converge to a stable traffic distribution, or would it devolve to perpetual congestion as users ping-pong between different routes (like commuters changing lanes in a traffic jam)? Note that whether the process converges at all pertains to the stability of the joint learning procedure, while whether it converges to Nash equilibria pertains to the rationality thereof: if the learning procedure converges to a non-Nash equilibrium, then each can do better by not following that learning procedure.

**Related work.** Despite the seeming simplicity, the existing literature has only provided scarce and qualitative answers to this question. This is in part due to the strong convergence mode conveyed by the question: while a large literature exists on this topic, much of them focuses on time-average convergence (i.e. convergence of the time average of the joint action), rather than last-iterate convergence (i.e. convergence of the joint action). However, not only is last-iterate convergence theoretically stronger and more appealing, it is also the only type of convergence that actually describes the system's evolution. This was a well-known point that was only recently rigorously illustrated in Mertikopoulos et al. (2018b), where it is shown that even though follow-the-regularized-leader (another no-regret learning algorithm) converges to a Nash equilibrium in linear zero-sum games in the sense of time-averages, actual joint action orbits Nash equilibria in perpetuity. Motivated by this consideration, a growing literature (Krichene et al., 2015; Lam et al., 2016; Zhou et al., 2017c; Palaiopanos et al., 2017; Zhou et al., 2017b; Mertikopoulos et al., 2017; Zhou et al., 2017a; 2020a; 2018; Zhou et al.; Mertikopoulos & Zhou, 2019) has devoted the efforts to obtaining last-iterate convergence results. However, due to the challenging nature of the problem, all of those lat-iterate convergence results are qualitative. In particular, except in strongly monotone games (Zhou et al. (2020b) very recently established a $O(1/T)$ last-iterate convergence rate for OGD learning with noisy feedback[1] in strongly monotone games), there are no quantitative, finite-time last-iterate convergence rates available[2].

Additionally, an important element in multi-agent online learning is that the horizon of play is typically unknown. As a result, no-regret learning algorithms need to be employed with a decreasing learning rate (e.g., because of a doubling trick or as a result of an explicit $O(1/t^{\alpha})$ step-size tuning). In particular, in order to achieve last-iterate convergence-to-Nash results, all of the above mentioned work rest crucially on using decreasing step-size (often converging to $0$ no slower than a particular rate) in their algorithm designs. This, however, leads to the following general tenet: *New information is utilized with decreasing weights*

From a rationality point of view, this is not only counter-intuitive – it flies at the face of established economic wisdom. Instead of *discounting* past information, players end up indirectly *reinforcing* it by assigning negligible weight to recent observations relative to those in the distant past. This negative recency bias is unjustifiable from economic micro-foundations and principles, and it cannot reasonably account for any plausible model of human/consumer behavior. The above naturally raises another important open question, one that, if answered, can bridge the gap between online learning and rationalizable economic micro-foundations: *Is no-regret learning without discounting recent information compatible with Nash equilibria?*

**Contributions.** Reflecting on those two gaps simultaneously, we are thus led to the following ambitious research question, one that aims to close two open questions at once: *Can we obtain finite-time last-iterate convergence rate using only non-decreasing step-size?* Our goal here is to make initial but significant progress in answering this question; our contributions are threefold.

First, we introduce a class of games that we call *cocoercive* and which contain all strongly monotone games as a special case. We show that if each player adopts OGD, then the joint action sequence converges in last-iterate to the set of Nash equilibria at a rate of $o(1/T)$. The convergence speed more specifically refers to how fast the gradient norm squared converges to $0$: note that in cocoercive games, gradient norm converges to $0$ if and only if the iterate converges to the set of Nash equilibria. To the best of our knowledge, this is the first rate that provides finite-time last-iterate convergence that moves beyond the strong monotonicity assumption.

Second, we study in depth the stochastic gradient feedback case, where each player adopts OGD in $\lambda$-cocoercive games, with gradient corrupted by a zero-mean, martingale-difference noise, whose variance is proportional to current gradient norm squared as assumed in the relative random noise model (Polyak, 1987). In this more challenging setting, we first establish that the joint action sequence converges in last-iterate to Nash equilibria almost surely under a constant step-size. The previous best such qualitative convergence is due to Mertikopoulos & Zhou (2019), which shows that such almost sure convergence is guaranteed in a variationally stable game. Despite the fact that variationally stable games contain cocoercive games as a subclass, our result is not covered by theirs because Mertikopoulos & Zhou (2019) assumes compact action set, where we con-

[1]The perfect gradient feedback case has a last-iterate convergence of $O(\rho^T)$, for some $0 < \rho < 1$ when the game is further Lipchitz. This follows from a classical result in variational inequality (Facchinei & Pang, 2007)

[2]Except in convex potential games. In that case, the problem of converging to Nash equilbiria reduces to a convex optimization problem, where standard techniques apply

sider unconstrained action set–a more challenging scenario since the action iterates can *a priori* be unbounded. Our result is further unique in that constant step-size is sufficient to achieve last-iterate almost sure convergence, while Mertikopoulos & Zhou (2019) requires decreasing step-size (that is square-summable-but-not-summable). Note that the relative random noise model is necessary for obtaining such constant step-size result: in an absolute random noise model (where the noise's second moment is bounded by a constant), the gradient descent iterate forms an ergodic and irreducible Markov chains, which induces an invariant measure that is supported on the entire action set, thereby making it impossible to obtain any convergence-to-Nash result. We then proceed a step further and characterize finite-time convergence rate. We establish two rates here: first, the expected time-average convergence rate is $O(1/T)$; second the expected last-iterate convergence rate is $O(a(T))$, where $a(T)$ depends on how fast the relative noise proportional constants decrease to 0. As a simple example, if those constants decrease to 0 at an $O(1/\sqrt{t})$ rate, then the last-iterate convergence rate is $O(1/\sqrt{T})$. For completeness (but due to space limitation), we also present in the appendix a parallel set of results–last-iterate almost sure convergence, time-average convergence rate and last-iterate convergence rate–for the absolute random noise model (under diminishing step-sizes of course).

Third, and even more surprisingly, we provide–to the best of our knowledge–the first adaptive gradient descent algorithm that has last-iterate convergence guarantees on games. In particular, the online gradient descent algorithms mentioned above–both in the deterministic and stochastic gradient case–requires the cocoercive constant $\lambda$ to be known beforehand. Thus, this calls for adaptive variants that do not require such knowledge. In the deterministic setting, we design an adaptive gradient descent algorithm that operates without needing to know $\lambda$ and adaptively chooses its step-size based on past gradients. We then show that the same $o(\frac{1}{T})$ last-iterate convergence rate can be achieved with non-decreasing step-size. Previously, the closest existing result is Bach & Levy (2019), which provided an adaptive algorithm on variational inequality with time-average convergence guarantees. However, providing adaptive algorithms for last-iterate convergence is much more challenging and Bach & Levy (2019) further requires the knowledge of the diameter of domain set (which they assume to be compact) in their adaptive algorithm, whereas we operate in unbounded domains. Our analysis relies on a novel double stopping time analysis, where the first stopping time characterizes the first time until gradient norm starts to monotonically decrease, and the second stopping time, after the first stopping has occured, characterizes the first time the underlying pseudo-contraction mapping starts to rapidly converge. We also provide the adaptive algorithm in the stochastic gradient feedback setting and establish the same finite-time last-iterate convergence guarantee. Note that our results only imply convergence in unconstrained strongly monotone games. Constrained coercive games is another interesting setting that would require a different set of techniques and further exploration.

## 2. Problem Setup

In this section, we present the definitions of a game with continuous action sets, which serves as a stage game and provides a reward function for each player in an online learning process. The key notion defined here is called $\lambda$-cococercivity, which is weaker than $\lambda$-strong monotonicity and covers a wider range of games.

### 2.1. Basic Definition and Notation

Throughout this paper, we focus on games played by a finite set of *players* $i \in \mathcal{N} = \{1, 2, \ldots, N\}$. During the learning process, each player selects an *action* $\mathbf{x}_i$ from a convex subset $\mathcal{X}_i$ of a finite-dimensional vector space $\mathbb{R}^{n_i}$ and their reward is determined by the profile $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ of all players' actions. Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm (in the corresponding vector space): other norms can be easily accomodated in our framework of course (and different $\mathcal{X}_i$'s can in general have different norms), although we will not bother with all of this since we do not plan to play with (and benefit from) complicated geometries.

**Definition 2.1** *A continuous game is a tuple $\mathcal{G} = (\mathcal{N}, \mathcal{X} = \prod_{i=1}^{N} \mathcal{X}_i, \{u_i\}_{i=1}^{N})$, where $\mathcal{N}$ is the set of $N$ players $\{1, 2, \ldots, N\}$, $\mathcal{X}_i$ is a convex set of some finite-dimensional vector space $\mathbb{R}^{n_i}$ representing the action space of player $i$, and $u_i : \mathcal{X} \to \mathbb{R}$ is the $i$-th player's payoff function satisfying:*

1. *For each $i \in \mathcal{N}$, the function $u_i(\mathbf{x})$ is continuous in $\mathbf{x}$.*
2. *For each $i \in \mathcal{N}$, the function $u_i$ is continuously differentiable in $\mathbf{x}_i$ and the partial gradient $\nabla_{\mathbf{x}_i} u_i(\mathbf{x})$ is Lipschitz continuous in $\mathbf{x}$.*

The notation $\mathbf{x}_{-i}$ denotes the joint action of all players but player $i$. Consequently, the joint action $\mathbf{x}$ will frequently be written as $(x_i; \mathbf{x}_{-i})$. Two important quantities are specified as follows:

**Definition 2.2** *$\mathbf{v}(\mathbf{x})$ is the profile of the players' individual payoff gradients, i.e., $\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), v_2(\mathbf{x}), \ldots, v_N(\mathbf{x}))$, where $v_i(\mathbf{x}) \triangleq \nabla_{x_i} u_i(\mathbf{x})$.*

We are looking at pure-Nash equilibria, because we are studying continuous games, where the action set is already

a finite-dimensional vector space, rather than a finite set as in the simpler finite games in which each players mixed strategy is a vector of probabilities in the simplex. In our setting, each action already lives in a continuum and we follow the standard definition of a pure Nash equilibrium.

**Definition 2.3** *$\mathbf{x}^* \in \mathcal{X}$ is called a (pure-strategy) Nash equilibrium of a game $\mathcal{G}$ if for each player $i \in \mathcal{N}$, it holds true that $u_i(x_i^*, \mathbf{x}_{-i}^*) \geq u_i(x_i, \mathbf{x}_{-i}^*)$ for each $x_i \in \mathcal{X}_i$.*

**Proposition 2.1** *In a continuous game $\mathcal{G}$, if $\mathbf{x}^* \in \mathcal{X}$ is a Nash equilibrium, then $(\mathbf{x} - \mathbf{x}^*)^\top \mathbf{v}(\mathbf{x}^*) \leq 0$ for all $\mathbf{x} \in \mathcal{X}$. The converse also holds true if the game is concave: for each $i \in \mathcal{N}$, the function $u_i(x_i; \mathbf{x}_{-i})$ is concave in $x_i$ for all $\mathbf{x}_{-i} \in \prod_{j \neq i} \mathcal{X}_j$.*

Proposition 2.1 is a classical result (see also (Mertikopoulos & Zhou, 2019) for a proof) and shows that the Nash equilibria of a concave game are precisely the solutions of the variational inequality $(\mathbf{x} - \mathbf{x}^*)^\top \mathbf{v}(\mathbf{x}^*) \leq 0$ for all $\mathbf{x} \in \mathcal{X}$.

**Definition 2.4** *A continuous game $\mathcal{G}$ is:*

1. ***montone*** *if $(\mathbf{x}' - \mathbf{x})^\top (\mathbf{v}(\mathbf{x}') - \mathbf{v}(\mathbf{x})) \leq 0$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$.*
2. ***strictly monotone*** *if $(\mathbf{x}' - \mathbf{x})^\top (\mathbf{v}(\mathbf{x}') - \mathbf{v}(\mathbf{x})) \leq 0$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$, with equality if and only if $\mathbf{x} = \mathbf{x}'$.*

**Remark 2.2** *We briefly highlight a few well-known properties of monotone and strictly monotone games. Monotone games are automatically concave games: a concave game (and hence a monotone game) is guaranteed to have a Nash equilibrium when all the action sets $\mathcal{X}_i$ are convex and compact. Otherwise, particularly in the unconstrained setting (each $\mathcal{X}_i = \mathbb{R}^{n_i}$), a Nash equilibrium may not exist. Many results on monotone games can be read off from the variational inequality literature (Facchinei & Pang, 2007); see all (Mertikopoulos & Zhou, 2019) for a detailed discussion.*

*In a strictly monotone game (first introduced in Rosen (1965) and referred to as diagonal strict concave games there), at most one Nash equilibrium exists; hence when all action sets are convex and compact, a strictly monotone game admits a unique Nash equilibrium. Additionally, when $\mathbf{v} = \nabla f$ for some (smooth) function $f$, $f$ is strictly concave. The notion strictly refers to the only if requirement in the condition. Many useful results regarding strictly monotone games (under convex and compact action sets) can be found in Rosen (1965).*

Proceeding a step further, we can define strongly monotone games:

**Definition 2.5** *A continuous game $\mathcal{G}$ is called $\lambda$-strongly monotone if the payoff strongly monotone condition holds: $(\mathbf{x}' - \mathbf{x})^\top (\mathbf{v}(\mathbf{x}') - \mathbf{v}(\mathbf{x})) \leq -\lambda \|\mathbf{x}' - \mathbf{x}\|^2$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$.*

Note that strongly monotone games are a subclass of strictly monotone games. One appealing feature of strongly monotone games is that the finite-time convergence rate can be derived in terms of $\|\mathbf{x}_t - \mathbf{x}^*\|$, where $\mathbf{x}^*$ is the unique Nash equilibrium (Zhou et al., 2020b) (under convex and compact action sets). On the other hand, $\mathbf{x}_t$ can possibly converge to a limit cycle or repeatedly hit the boundary in monotone games (Mertikopoulos et al., 2018a; Daskalakis et al., 2018) despite that the time-average $(\sum_{j=1}^t \mathbf{x}_j)/t$ converges. More recently, Mertikopoulos & Zhou (2019) analyzed online mirror descent (OMD) learning (which contains OGD as a special case) in variational stable games (under convex and compact action sets) and proved that last-iterate convergence to Nash equilibria holds almost surely in the presence of imperfect feedback (i.e. gradient corrupted by an unbiased noise). This result is surprising since the notion of variational stablitly is much weaker than strict monotonicity (hence qualitative convergence to Nash in strictly monotone games is guaranteed), showing that strong monotonicity is unnecessary for last-iterate convergence of OGD learning.

However, there are no last-iterate convergence rates available for strictly monotone games (unconstrained or constrained), and such a result does not seem possible because the strictness gap can be made arbitrarily small (and yielding arbitrarily slow rates). In fact, without the quadratic growth of strong monotonicity, it seems impossible to attain the rate for $\|\mathbf{x}_t - \mathbf{x}^*\|$ and, moreover, using the method with a constant step-size is completely out of reach of the techniques of Zhou et al. (2020b). Finally, we remark that fully adaptive and parameter-free learning methods are also missing from game-theoretic analyses to date.

### 2.2. $\lambda$-Cococercive Games

**Definition 2.6** *A continuous game $\mathcal{G}$ is called $\lambda$-cocoercive if the payoff cocoercive condition holds: $(\mathbf{x}' - \mathbf{x})^\top (\mathbf{v}(\mathbf{x}') - \mathbf{v}(\mathbf{x})) \leq -\lambda \|\mathbf{v}(\mathbf{x}') - \mathbf{v}(\mathbf{x})\|^2$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$.*

In this paper, we focus on the unconstrained $\lambda$-cocoercive game in which $\mathcal{X}_i = \mathbb{R}^{n_i}$, while the existing literature invariably assumes compactness, which is the constrained setting. The analysis for these two settings should be considered as complementary (and results in Zhou et al. (2020b) would not apply to unconstrained strongly monotone cases). We study the unconstrained setting (which is another indication that new techniques different from previous work are needed) because the adaptive algorithms and analyses are easier to present. Three important comments are in order.

First, a cocoercive game is a monotone game (as can be easily seen from the definitions), but cocoercive games neither contain nor belong to strictly monotone games. When a Nash equilibrium exists in a cocoercive game, it may not be unique; further, all Nash equilibria of a cococercive

game shares the same individual payoff gradient (either constrained or unconstrained): neither of these two properties hold in a strictly monotone game. For a simple one-player example where the cost function $f(x) = x^2$ for $x < 0$ and 0 otherwise: this game is cocoercive but not strictly monotone. Moreover, $(\mathbf{x}' - \mathbf{x})^\top(\mathbf{v}(\mathbf{x}') - \mathbf{v}(\mathbf{x})) = 0$ only implies that $\mathbf{v}(\mathbf{x}') = \mathbf{v}(\mathbf{x})$ and $\mathbf{x}' = \mathbf{x}$ does not necessarily hold true.

Second, the unconstrained cocoercive games may not always have a Nash equilibrium since we lifted the compactness assumption. Accordingly, all of our subsequent convergence results are stated for games that do have Nash equilibria: we did so because we want our results to apply to all cocoercive games that have Nash equilibria. That said, many additional sufficient conditions can be imposed on a cocoercive game to ensure the existence of a Nash equilrium. One such sufficient condition is the coercivity of the costs: the costs go to infinity as joint actions go to infinity (as already alluded to, we didnt assume both cocoercivity and coercivity because that would eliminate other cocoercive games that admit Nash equilibria, which is more restrictive).

Thirdly, we remark that it is more difficult to analyze the convergence property of online algorithms in unconstrained setting, especially when the feedback information is noisy, since the iterates are not necessarily assumed to be bounded. Further, since in a cocoercive game, $\mathbf{x}^* \in \mathcal{X}^*$ is a Nash equilibrium if and only if $\mathbf{v}(\mathbf{x}^*) = 0$, the natural candidate for measuring convergence (i.e. optimality gap) is $\epsilon(\mathbf{x}) = \|\mathbf{v}(\mathbf{x})\|^2$.

### 2.3. Learning via Online Gradient Descent

We describe the *online gradient descent* (OGD) algorithm in our game-theoretical setting. Intuitively, the main idea is: At each stage, every player $i \in \mathcal{N}$ gets an estimate $\hat{v}_i$ of the individual gradient of their payoff function at current action profile, possibly subject to noise and uncertainty. Subsequently, they choose an action $x_i$ for the next stage using the current action and feedback $\hat{v}_i$, and continue playing.

Formally, starting with some arbitrarily (and possibly uninformed) iterate $\mathbf{x}_0 \in \mathbb{R}^n$ at $t = 0$, the scheme can be described via the recursion

$$x_{i,t+1} \;=\; x_{i,t} + \eta_{t+1}\hat{v}_{i,t+1}, \tag{1}$$

where $t \geq 0$ denotes the stage of process, $\hat{v}_{i,t}$ is an estimate of the individual payoff gradient $v_i(\mathbf{x}_t)$ of player $i$ at stage $t$. The learning rate $\eta_t > 0$ is a nonincreasing sequence which can be of the form $c/t^p$ for some $p \in [0, 1]$.

**Feedback and uncertainty:** we assume that each player $i \in \mathcal{N}$ has access to a "black box" feedback mechanism – an *oracle* – which returns an estimate of their payoff gradients at their current action profile. This information can be imperfect for a multitude of reasons; see Mertikopoulos & Zhou (2019, Section 3.1). With all this in mind, we consider the following noisy feedback model:

$$\hat{v}_{i,t+1} \;=\; v_i(\mathbf{x}_t) + \xi_{i,t+1}, \tag{2}$$

where the noise process $\xi_t = (\xi_{i,t})_{i\in\mathcal{N}}$ is an $L^2$-bounded martingale difference adapted to the history $(\mathcal{F}_t)_{t\geq 1}$ of $\mathbf{x}_t$ (i.e., $\xi_t$ is $\mathcal{F}_t$-measurable but $\xi_{t+1}$ isn't).

We focus on two types of random noise proposed by (Polyak, 1987). The first type is called **relative random noise**:

$$\mathbb{E}[\xi_{t+1} \mid \mathcal{F}_t] = 0, \quad \mathbb{E}[\|\xi_{t+1}\|^2 \mid \mathcal{F}_t] \leq \tau_t\|\mathbf{v}(\mathbf{x}_t)\|^2. \tag{3}$$

and the second type is called **absolute random noise**:

$$\mathbb{E}[\xi_{t+1} \mid \mathcal{F}_t] = 0, \quad \mathbb{E}[\|\xi_{t+1}\|^2 \mid \mathcal{F}_t] \leq \sigma_t^2, \tag{4}$$

The above condition is mild (the i.i.d. condition is not imposed) and allows for a broad range of error processes. For the relative random noise, the variance decreases as it approaches a Nash equilibrium which admits better convergence rate of learning algorithms.

## 3. Convergence under Perfect Feedback

In this section, we analyze the convergence property of OGD learning under perfect feedback. In particular, we show that the finite-time last-iterate convergence rate is $o(1/T)$ regardless of fully adaptive learning rates. To our knowledge, the proof techniques for analyzing adaptive OGD learning is new and can be of independent interests.

### 3.1. OGD Learning

We first provide a lemma which shows that $\|\mathbf{v}(\mathbf{x}_t)\|^2$ is nonnegative, nonincreasing and summable.

**Lemma 3.1** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. Under the condition that $\eta_t = \eta \in (0, \lambda]$, the OGD iterate $\mathbf{x}_t$ satisfies for all $t \geq 0$ that $\|\mathbf{v}(\mathbf{x}_{t+1})\| \leq \|\mathbf{v}(\mathbf{x}_t)\|$ and*

$$\begin{aligned} \|\mathbf{x}_t - \Pi_{\mathcal{X}^*}(\mathbf{x}_0)\| &\leq \|\mathbf{x}_0 - \Pi_{\mathcal{X}^*}(\mathbf{x}_0)\|, \\ \sum_{t=0}^{+\infty} \|\mathbf{v}(\mathbf{x}_t)\|^2 &\leq \frac{\|\mathbf{x}_0 - \Pi_{\mathcal{X}^*}(\mathbf{x}_0)\|^2}{\eta\lambda}. \end{aligned}$$

**Remark 3.2** *For the OGD learning with perfect feedback and constant step-size, the update formula in Eq. (1) implies that $\|\mathbf{v}(\mathbf{x}_t)\|^2 = \|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2/\eta^2$. This implies that $\|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2$ also serves as the candidate for an optimality gap function. Such quantity is called the iterative gap and frequently used to construct the stopping criteria in practice.*

Now we are ready to present our main results on the last-iterate convergence rate of OGD learning.

**Algorithm 1** Adaptive Online Gradient Descent

**Initialization:** $\mathbf{x}_0 \in \mathbb{R}^n$, $\eta_1 = 1/\beta_1$ for some $\beta_1 > 0$ and tuning parameter $r > 1$.
**for** $t = 0, 1, 2, \ldots$ **do**
  **for** $i = 1, 2, \ldots, N$ **do**
    $x_i^{t+1} \leftarrow x_i + \eta_{t+1}\mathbf{v}(\mathbf{x}_t)$.
    **if** $\|\mathbf{v}(\mathbf{x}_{t+1})\| > \|\mathbf{v}(\mathbf{x}_t)\|$ **then**
      $\beta_{t+2} \leftarrow r\beta_{t+1}$.
    **else**
      $\beta_{t+2} \leftarrow \beta_{t+1}$.
    **end if**
    $\eta_{t+2} \leftarrow 1/\sqrt{\beta_{t+2} + \sum_{j=0}^{t} \|\mathbf{v}(\mathbf{x}_j)\|^2}$.
  **end for**
**end for**

**Theorem 3.3** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. Under the condition that $\eta_t = \eta \in (0, \lambda]$, the OGD iterate $\mathbf{x}_t$ satisfies that $\epsilon(\mathbf{x}_T) = o(1/T)$.*

*Proof.* Lemma 3.1 implies that $\{\|\mathbf{v}(\mathbf{x}_t)\|^2\}_{t\geq 0}$ is nonnegative, nonincreasing and $\sum_{t=0}^{+\infty} \|\mathbf{v}(\mathbf{x}_t)\|^2 < +\infty$. Therefore,

$$T\|\mathbf{v}(\mathbf{x}_{2T-1})\|^2 \leq \sum_{t=T}^{2T-1} \|\mathbf{v}(\mathbf{x}_t)\|^2 \;\to\; 0 \text{ as } T \to +\infty.$$

which implies that $\|\mathbf{v}(\mathbf{x}_T)\|^2 = o(1/T)$. By the definition of $\epsilon(\mathbf{x})$, we conclude the desired result. □

### 3.2. Adaptive OGD Learning

Our main results in this subsection is the last-iterate convergence rate of Algorithm 1. Here the algorithm requires no prior knowledge of $\lambda$ but still achieves the rate of $o(1/T)$. To facilitate the readers, we summarize the results in the following theorem and provide the detailed proof.

**Theorem 3.4** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. The adaptive OGD iterate $\mathbf{x}_t$ satisfies that $\epsilon(\mathbf{x}_T) = o(1/T)$.*

*Proof.* Since the step-size sequence $\{\eta_t\}_{t\geq 1}$ is nonincreasing, we define the first iconic time in our analysis as follows,

$$t^* = \max\{t \geq 0 \mid \eta_{t+1} > \lambda\}.$$

In what follows, we prove the last-iterate convergence rate for two cases: $t^* = +\infty$ (**Case I**) and $t^* < +\infty$ (**Case II**).

**Case I.** First, we have $1/\lambda^2 - \beta_0 \geq 0$ since $\eta_0 > \lambda$. Note that $\beta_{t+2} \leftarrow r\beta_{t+1}$ with $r > 1$ is updated when $\|\mathbf{v}(\mathbf{x}_{t+1})\| > \|\mathbf{v}(\mathbf{x}_t)\|$, there exists $T_0 > 0$ such that $\|\mathbf{v}(\mathbf{x}_{t+1})\| \leq \|\mathbf{v}(\mathbf{x}_t)\|$ for all $t \geq T_0$. If not, then $\beta_t \to +\infty$ as $t \to +\infty$ and $\eta_t \to 0$. However, $t^* = +\infty$ implies that $\eta_{t+1} \geq \lambda$ for all $t \geq 1$. This leads to a contradiction. Furthermore, it holds true for all $t \geq 0$ that

$$\sum_{j=0}^{t} \|\mathbf{v}(\mathbf{x}_j)\|^2 \;\leq\; \frac{1}{\lambda^2} - \beta_{t+2} \;\leq\; \frac{1}{\lambda^2} - \beta_0 \;<\; +\infty.$$

By starting the sequence at a later index $T_0$, we have $\sum_{t\geq T_0} \|\mathbf{v}(\mathbf{x}_t)\|^2 < +\infty$ and $\|\mathbf{v}(\mathbf{x}_{t+1})\| \leq \|\mathbf{v}(\mathbf{x}_t)\|$ for all $t \geq T_0$. Using the same argument as in Theorem 3.3, the adaptive OGD iterate $\mathbf{x}_t$ satisfies that $\epsilon(\mathbf{x}_T) = o(1/T)$.

**Case II.** First, we claim that $\|\mathbf{x}_t - \Pi_{\mathcal{X}^*}(\mathbf{x}_{t^*})\| \leq D$ where $D = \max_{1\leq t\leq t^*} \|\mathbf{x}_t - \Pi_{\mathcal{X}^*}(\mathbf{x}_{t^*})\|$. Indeed, it suffices to show that $\|\mathbf{x}_t - \Pi_{\mathcal{X}^*}(\mathbf{x}_{t^*})\| \leq \|\mathbf{x}_{t^*} - \Pi_{\mathcal{X}^*}(\mathbf{x}_{t^*})\|$ holds for $t > t^*$. By the definition of $t^*$, we have $\eta_{t+1} \leq \lambda$ for all $t > t^*$. The desired inequality follows from Lemma 3.1.

Using the update formula (cf. Eq. (1)), we have

$$\begin{aligned}(\mathbf{x}_{t+1} - \mathbf{x}^*)^\top \mathbf{v}(\mathbf{x}_t) \;&=\; \frac{1}{2\eta_{t+1}} \big(\|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2 \\ &\quad + \|\mathbf{x}^* - \mathbf{x}_{t+1}\|^2 - \|\mathbf{x}^* - \mathbf{x}_t\|^2\big).\end{aligned}$$

Using the update formula of OGD learning, we have

$$\begin{aligned}\lambda\|\mathbf{v}(\mathbf{x}_t)\|^2 \;&\leq\; \frac{1}{\eta_{t+1}} \left(\|\mathbf{x}^* - \mathbf{x}_t\|^2 - \|\mathbf{x}^* - \mathbf{x}_{t+1}\|^2\right) \\ &\quad + \left(\frac{1}{\lambda} - \frac{1}{\eta_{t+1}}\right) \|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2. \qquad (5)\end{aligned}$$

Summing up Eq. (5) over $t = 0, 1, 2, \ldots, T$ yields that

$$\begin{aligned}\sum_{t=0}^{T} \lambda\|\mathbf{v}(\mathbf{x}_t)\|^2 \;&\leq\; \sum_{t=1}^{T} \|\mathbf{x}^* - \mathbf{x}_t\|^2 \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right) \\ &\quad + \frac{\|\mathbf{x}^* - \mathbf{x}_0\|^2}{\eta_1} + \sum_{t=0}^{T} \left(\frac{1}{\lambda} - \frac{1}{\eta_{t+1}}\right) \|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2.\end{aligned}$$

Since the step-size sequence $\{\eta_t\}_{t\geq 1}$ is nonincreasing, we have $1/\eta_{t+1} \geq 1/\eta_t$. Letting $\mathbf{x}^* = \Pi_{\mathcal{X}^*}(\mathbf{x}_{t^*})$, we notice that $\|\mathbf{x}^* - \mathbf{x}_t\| \leq D$ for all $0 \leq t \leq T$. Putting these pieces together yields that

$$\sum_{t=0}^{T} \lambda\|\mathbf{v}(\mathbf{x}_t)\|^2 \;\leq\; \frac{D^2}{\eta_{T+1}} + \sum_{t=0}^{T} \left(\frac{1}{\lambda} - \frac{1}{\eta_{t+1}}\right) \|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2.$$

To proceed, we define the second iconic time as

$$t_1^* \;=\; \max\left\{t \geq 0 \mid \eta_{t+1} > \frac{\lambda}{2D^2 + 1}\right\} \;>\; t^*.$$

Suppose that $t_1^* = +\infty$, it is straightforward to show that the adaptive OGD iterate $\mathbf{x}_t$ satisfies that $\epsilon(\mathbf{x}_T) = o(1/T)$ using the same argument in **Case I**.

Next, we consider $t_1^* < +\infty$. Indeed, we recall that $\eta_{t+1} \leq \lambda$ for all $t > t_1^*$ which implies that $1/\lambda - 1/\eta_{t+1} \leq 0$. Since $\|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2 = \eta_{t+1}^2 \|\mathbf{v}(\mathbf{x}_t)\|^2$ (cf. Eq. (1)) and assume $T$ sufficiently large without loss of generality, we have

$$\sum_{t=0}^{T} \lambda \|\mathbf{v}(\mathbf{x}_t)\|^2 \;\leq\; \frac{D^2}{\eta_{T+1}} + \sum_{t=0}^{t_1^*} \frac{\eta_{t+1}^2 \|\mathbf{v}(\mathbf{x}_t)\|^2}{\lambda} \;=\; \mathrm{I} + \mathrm{II}.$$

Before bounding term I and II, we present two technical lemmas which is crucial to our subsequent analysis; see (Bach & Levy, 2019, Lemma A.1 and A.2) for the detailed proof.

**Lemma 3.5** *For a sequence of numbers $a_0, a_1, \ldots, a_n \in [0, a]$ and $b \geq 0$, the following inequality holds:*

$$\sqrt{b + \sum_{i=0}^{n-1} a_i} - \sqrt{b} \;\leq\; \sum_{i=0}^{n} \frac{a_i}{\sqrt{b + \sum_{j=0}^{i-1} a_j}} \leq \frac{2a}{\sqrt{b}} + 3\sqrt{a} + 3\sqrt{b + \sum_{i=0}^{n-1} a_i}.$$

**Lemma 3.6** *For a sequence of numbers $a_0, a_1, \ldots, a_n \in [0, a]$ and $b \geq 0$, the following inequality holds:*

$$\sum_{i=0}^{n} \frac{a_i}{b + \sum_{j=0}^{i-1} a_j} \leq 2 + \frac{4a}{b} + 2\log\left(1 + \sum_{i=0}^{n-1} \frac{a_i}{b}\right).$$

**Bounding term I:** By the definition of $t_1^*$ and Lemma 3.1, we have $\beta_t = \beta_{t_1^*+1}$ for all $t > t_1^*$. Thus, we derive from the definition of $\eta_t$ that

$$\mathrm{I} \leq D^2 \sqrt{\beta_{T+1} + \sum_{j=0}^{T-1} \|\mathbf{v}(\mathbf{x}_j)\|^2} \leq D^2 \sqrt{\beta_{t_1^*+1} + \sum_{j=0}^{T-1} \|\mathbf{v}(\mathbf{x}_j)\|^2}.$$

Since $\|\mathbf{x}_t - \Pi_{\mathcal{X}^*}(\mathbf{x}_{t^*})\| \leq D$ for all $t \geq 0$. Since the notion of $\lambda$-cocercivity implies the notion of $(1/\lambda)$-Lipschiz continuity, we have

$$\|\mathbf{v}(\mathbf{x}_t)\|^2 \;\leq\; \frac{\|\mathbf{x}_t - \Pi_{\mathcal{X}^*}(\mathbf{x}_{t^*})\|^2}{\lambda^2} \;\leq\; \frac{D^2}{\lambda^2}. \tag{6}$$

Using the first inequality in Lemma 3.5, we have

$$\begin{aligned}
\mathrm{I} \;&\leq\; D^2 \sqrt{\beta_{t_1^*+1}} + \sum_{t=0}^{T} \frac{D^2 \|\mathbf{v}(\mathbf{x}_t)\|^2}{\sqrt{\beta_{t_1^*+1} + \sum_{j=0}^{t-1} \|\mathbf{v}(\mathbf{x}_j)\|^2}} \\
&\leq\; D^2 \sqrt{\beta_{t_1^*+1}} + \sum_{t=0}^{t_1^*} \frac{D^2 \|\mathbf{v}(\mathbf{x}_t)\|^2}{\sqrt{\beta_{t_1^*+1} + \sum_{j=0}^{t-1} \|\mathbf{v}(\mathbf{x}_j)\|^2}} \\
&\quad + \sum_{t=t_1^*+1}^{T} D^2 \eta_{t+1} \|\mathbf{v}(\mathbf{x}_t)\|^2.
\end{aligned} \tag{7}$$

Since $\eta_{t+1} \leq \lambda / 2D^2$ for all $t > t_1^*$, we have

$$\sum_{t=t_1^*+1}^{T} D^2 \eta_{t+1} \|\mathbf{v}(\mathbf{x}_t)\|^2 \;\leq\; \sum_{t=t_1^*+1}^{T} \frac{\lambda \|\mathbf{v}(\mathbf{x}_t)\|^2}{2}. \tag{8}$$

Using the second inequality in Lemma 3.5,

$$\begin{aligned}
&\sum_{t=0}^{t_1^*} \frac{D^2 \|\mathbf{v}(\mathbf{x}_t)\|^2}{\sqrt{\beta_{t_1^*+1} + \sum_{j=0}^{t-1} \|\mathbf{v}(\mathbf{x}_j)\|^2}} \\
\leq\; &\frac{2D^2}{\lambda^2 \sqrt{\beta_{t_1^*+1}}} + \frac{3D}{\lambda} + 3\sqrt{\beta_{t_1^*+1} + \sum_{j=0}^{t_1^*-1} \|\mathbf{v}(\mathbf{x}_j)\|^2}.
\end{aligned} \tag{9}$$

By the definition of $\eta_t$, we have

$$\sqrt{\beta_{t_1^*+1} + \sum_{j=0}^{t_1^*-1} \|\mathbf{v}(\mathbf{x}_j)\|^2} \;=\; \frac{1}{\eta_{t_1^*+1}} \;<\; \frac{2D^2 + 1}{\lambda}. \tag{10}$$

Putting Eq. (7)-(10) together yields that

$$\begin{aligned}
\mathrm{I} \;&\leq\; D^2 \sqrt{\beta_{t_1^*+1}} + \frac{2D^2}{\lambda^2 \sqrt{\beta_{t_1^*+1}}} + \frac{3 + 3D + 6D^2}{\lambda} \\
&\quad + \sum_{t=t_1^*+1}^{T} \frac{\lambda \|\mathbf{v}(\mathbf{x}_t)\|^2}{2}.
\end{aligned}$$

**Bounding term II:** By the definition of $\eta_t$ and noting that $\beta_t \geq \beta_1$ for all $t \geq 1$, we have

$$\mathrm{II} \;\leq\; \frac{1}{\lambda} \left( \sum_{t=0}^{t_1^*} \frac{\|\mathbf{v}(\mathbf{x}_t)\|^2}{\beta_1 + \sum_{j=0}^{t-1} \|\mathbf{v}(\mathbf{x}_j)\|^2} \right)$$

Recalling Eq. (6), we can apply Lemma 3.6 with Eq. (10) to obtain that

$$\begin{aligned}
\mathrm{II} \;&\leq\; \frac{1}{\lambda} \left( 2 + \frac{4D^2}{\lambda^2 \beta_1} + 2\log\left( 1 + \frac{1}{\beta_1} \sum_{j=0}^{t_1^*-1} \|\mathbf{v}(\mathbf{x}_j)\|^2 \right) \right) \\
&\leq\; \frac{1}{\lambda} \left( 2 + \frac{4D^2}{\lambda^2 \beta_1} + 2\log\left( 1 + \frac{8D^4 + 2}{\lambda^2 \beta_1} \right) \right).
\end{aligned}$$

Therefore, we conclude that

$$\begin{aligned}
\sum_{t=0}^{T} \frac{\lambda \|\mathbf{v}(\mathbf{x}_t)\|^2}{2} \;&\leq\; D^2 \sqrt{\beta_{t_1^*+1}} + \frac{2D^2}{\lambda^2 \sqrt{\beta_{t_1^*+1}}} \\
&+ \frac{3 + 3D + 6D^2}{\lambda} + \frac{1}{\lambda} \left( 2 + \frac{4D^2}{\lambda^2 \beta_1} + 2\log\left( 1 + \frac{8D^4 + 2}{\lambda^2 \beta_1} \right) \right).
\end{aligned}$$

which implies that $\sum_{t=0}^{T} \|\mathbf{v}(\mathbf{x}_t)\|^2$ is bounded by a constant for all $T \geq 0$. By starting the sequence at a later index $t_1^*$, we have $\|\mathbf{v}(\mathbf{x}_{t+1})\| \leq \|\mathbf{v}(\mathbf{x}_t)\|$ for all $t \geq t_1^*$ and $\sum_{t \geq t_1^*} \|\mathbf{v}(\mathbf{x}_t)\|^2 < +\infty$. Using the same argument as in Theorem 3.3, we conclude that the adaptive OGD iterate $\mathbf{x}_t$ satisfies that $\epsilon(\mathbf{x}_T) = o(1/T)$. □

**Remark 3.7** *In our proof, $D > 0$ is the constant that depends on the set of Nash equilibrium, and in stating the bound this way, we followed the standard tradition in optimization where the bound (on either $f(\mathbf{x}_t) - f(\mathbf{x}^*)$ or $\|\nabla f(\mathbf{x}^t)\|^2$) would depend on $\|\mathbf{x}_0 - \mathbf{x}^*\|$ (a constant that cannot be avoided). Here, our bound similarly depends on this constant, except the game setting is more complicated (since there is no common objective) so this constant depends on the first few iterates as well (not just the initial iterate $\mathbf{x}^0$): more precisely, $D = \max_{1\le t\le t^*} \|\mathbf{x}_t - \Pi_{\mathcal{X}^*}(\mathbf{x}_{t^*})\|$.*

## 4. Convergence under Imperfect Feedback with Relative Random Noise

In this section, we analyze the convergence property of OGD learning under imperfect feedback with relative random noise (3). In particular, we show that the almost sure last-iterate convergence is guaranteed and the finite-time average-iterate convergence rate is $O(1/T)$ when $0 < \tau_t \le \tau < +\infty$. More importantly, we get the finite-time last-iterate convergence rate when $\tau_t$ satisfies certain summable condition (13).

### 4.1. Almost Sure Last-Iterate Convergence

In this subsection, we establish the almost sure last-iterate convergence under imperfect feedback with relative random noise. The appealing feature here is that the convergence results provably hold with a constant step-size. The first and second lemmas provide two different key inequalities for $\mathbf{x}_t$ and $\mathbb{E}[\epsilon(\mathbf{x}_t)]$ respectively.

**Lemma 4.1** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. Under the noisy model (2), the noisy OGD iterate $\mathbf{x}_t$ satisfies for any Nash equilibrium $\mathbf{x}^* \in \mathcal{X}^*$ that*

$$\begin{aligned}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 \;\le\; &\|\mathbf{x}_t - \mathbf{x}^*\|^2 + 2\eta_{t+1}^2\|\xi_{t+1}\|^2 \\ &-(2\lambda\eta_{t+1} - 2\eta_{t+1}^2)\|\mathbf{v}(\mathbf{x}_t)\|^2 + 2\eta_{t+1}(\mathbf{x}_t - \mathbf{x}^*)^\top\xi_{t+1}.\end{aligned} \tag{11}$$

**Lemma 4.2** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. Under the noisy model (2) with relative random noise (3) and the step-size sequence $\eta_t \in (0, \lambda)$. The noisy OGD iterate $\mathbf{x}_t$ satisfies $\mathbb{E}[\epsilon(\mathbf{x}_{t+1})] \le \mathbb{E}[\epsilon(\mathbf{x}_t)] + \tau_t\|\mathbf{v}(\mathbf{x}_t)\|^2/\lambda\eta_{t+1}$.*

Now we are ready to characterize the almost sure last-iterate convergence. Note that the condition imposed on $\tau_t$ is minimal and $\eta_t = \eta \in [\underline{\eta}, \overline{\eta}]$ is allowed for all $t \ge 1$.

**Theorem 4.3** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. Under the noisy model (2) with relative random noise (3) satisfying $\tau_t \in (0, \tau]$ for some $\tau < +\infty$ and the step-size sequence satisfying $0 < \underline{\eta} \le \eta_t \le \overline{\eta} < \lambda/(1+\tau)$ for all $t \ge 1$. The noisy OGD iterate $\mathbf{x}_t$ converges to $\mathcal{X}^*$ almost surely.*

*Proof.* We obtain the following inenquality by taking the expectation of both sides of Eq. (11) (cf. Lemma 4.1) conditioned on $\mathcal{F}_t$:

$$\begin{aligned}\mathbb{E}[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 \mid \mathcal{F}_t] \;\le\; &\|\mathbf{x}_t - \mathbf{x}^*\|^2 \\ &-(2\lambda\eta_{t+1} - 2\eta_{t+1}^2)\|\mathbf{v}(\mathbf{x}_t)\|^2 + 2\eta_{t+1}^2\mathbb{E}[\|\xi_{t+1}\|^2 \mid \mathcal{F}_t] \\ &+2\eta_{t+1}\mathbb{E}[(\mathbf{x}_t - \mathbf{x}^*)^\top\xi_{t+1} \mid \mathcal{F}_t].\end{aligned}$$

Since the noisy model (2) is with relative random noise (3) satisfying $\tau_t \in (0, \tau)$ for some $\tau < +\infty$, we have $\mathbb{E}[(\mathbf{x}_t - \mathbf{x}^*)^\top\xi_{t+1} \mid \mathcal{F}_t] = 0$ and $\mathbb{E}[\|\xi_{t+1}\|^2 \mid \mathcal{F}_t] \le \tau\|\mathbf{v}(\mathbf{x}_t)\|^2$. Therefore, we have

$$\begin{aligned}\mathbb{E}[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 \mid \mathcal{F}_t] \;\le\; &\|\mathbf{x}_t - \mathbf{x}^*\|^2 \\ &-2(\lambda - \overline{\eta} - \tau\overline{\eta})\eta_{t+1}\|\mathbf{v}(\mathbf{x}_t)\|^2.\end{aligned} \tag{12}$$

Since $\eta_t > 0$ and $\overline{\eta} < \lambda/(1+\tau)$, we let $M_t = \|\mathbf{x}_t - \mathbf{x}^*\|^2$ and obtain that $M_t$ is an nonnegative supermartingale. Then Doob's martingale convergence theorem shows that $M_n$ converges to an nonnegative and integrable random variable almost surely. Let $M_\infty = \lim_{t\to+\infty} M_t$, it suffices to show that $M_\infty = 0$ almost surely now. We assume to contrary that, there exists $m > 0$ such that $M_\infty > m$ with positive probability. Then $M_t > m/2$ for sufficiently large $t$ with positive probability. Formally, there exists $\delta > 0$ such that

$$\text{Prob}(M_t > m/2 \text{ for sufficiently large } t) \;\ge\; \delta.$$

By the definition of $M_t$ and recalling that $\mathbf{x}^* \in \mathcal{X}^*$ can be any Nash equilibrium, we let $U$ be a $(m/2)$-neighborhood of $\mathcal{X}^*$ and obtain that $\mathbf{x}_t \notin U$ for sufficiently large $t$ with positive probability. Since $\|\mathbf{v}(\mathbf{x})\| = 0$ if and only if $\mathbf{x} \in \mathcal{X}^*$, there exists $c > 0$ such that $\|\mathbf{v}(\mathbf{x})\| \ge c$ for sufficiently large $t$ with positive probability. Therefore, we conclude that $\mathbb{E}[\|\mathbf{v}(\mathbf{x}_t)\|^2] \not\to 0$ as $t \to +\infty$.

On the other hand, by taking the expectation of Eq. (20) and using the condition $\eta_t \ge \underline{\eta} > 0$ for all $t \ge 1$, we have

$$\mathbb{E}[\|\mathbf{v}(\mathbf{x}_t)\|^2] \;\le\; \frac{\mathbb{E}[\|\mathbf{x}_t - \mathbf{x}^*\|^2] - \mathbb{E}[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2]}{2(\lambda - \overline{\eta} - \tau\overline{\eta})\underline{\eta}}.$$

This implies that $\sum_{t=0}^\infty \mathbb{E}[\|\mathbf{v}(\mathbf{x}_t)\|^2] < +\infty$ and hence $\mathbb{E}[\|\mathbf{v}(\mathbf{x}_t)\|^2] \to 0$ as $t \to +\infty$ which contradicts the previous argument. This completes the proof. □

### 4.2. Finite-Time Convergence Rate: Time-Average and Last-Iterate

In this subsection, we focus on deriving two types of rates: the time-average and last-iterate convergence rates, as formalized by the following theorems.

**Theorem 4.4** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. Under the noisy model (2) with relative random noise (3) satisfying $\tau_t \in (0, \tau]$ for some $\tau < +\infty$ and a step-size sequence satisfying $0 < \underline{\eta} \le \eta_t \le \overline{\eta} < \lambda/(1+\tau)$ for all $t \ge 1$, the noisy iterate $\mathbf{x}_t$ satisfies $\frac{1}{T+1}(\mathbb{E}[\sum_{t=0}^T \epsilon(\mathbf{x}_t)]) = O(1/T)$.*

Inspired by Lemma 4.2, we impose an intuitive condition on the variance ratio of noisy process $\{\tau_t\}_{t\ge 0}$. More specifically, $\{\tau_t\}_{t\ge 0}$ is an nonincreasing sequence and there exists a function $a : \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $a(t) = o(1)$ such that

$$\frac{1}{T+1}\left(\sum_{t=0}^{T-1} \tau_t\right) \;=\; O(a(T)). \tag{13}$$

**Remark 4.5** *The condition (13) is fairly mild. Indeed, the decaying rate $a(t)$ can be very slow which still guarantees the finite-time last-iterate convergence rate. For some typical examples, we have $a(t) = \log\log(t)/t$ if $\tau_t = 1/t\log(t)$ and $a(t) = \log(t)/t$ if $\tau_t = 1/t$. When $\tau_t = \Omega(1/t)$, we have $a(t) = \tau_t$, such as $a(t) = 1/\sqrt{t}$ if $\tau_t = 1/\sqrt{t}$ and $a(t) = 1/\log\log(t)$ if $\tau_t = 1/\log\log(t)$. Under this condition, we can derive the last-iterate convergence rate given the decaying rate of $\tau_t$ as $t \to +\infty$.*

Under the condition (13), the finite-time last-iterate convergence rate can be derived under certain step-size sequences.

**Theorem 4.6** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. Under the noisy model (2) with relative random noise (3) satisfying Eq. (13) and the step-size sequence satisfying $0 < \underline{\eta} \le \eta_t \le \overline{\eta} < \lambda/(1+\tau)$ for all $t \ge 1$, the noisy OGD iterate $\mathbf{x}_t$ satisfies*

$$\mathbb{E}[\epsilon(\mathbf{x}_T)] \;=\; \begin{cases} O(a(T)) & \text{if } a(T) = \Omega(1/T), \\ O(1/T) & \text{otherwise.} \end{cases}$$

*Proof.* Using Lemma 4.2 and $\eta_t \ge \underline{\eta} > 0$, we have

$$\mathbb{E}[\epsilon(\mathbf{x}_T)] \;\le\; \mathbb{E}[\epsilon(\mathbf{x}_t)] + \frac{\sum_{j=t}^{T-1} \tau_j \|\mathbf{v}(\mathbf{x}_j)\|^2}{\lambda \underline{\eta}}.$$

Summing up the above inequality over $t = 0, \ldots, T$ yields

$$(T+1)\mathbb{E}[\epsilon(\mathbf{x}_T)] \le \sum_{t=0}^{T} \mathbb{E}[\epsilon(\mathbf{x}_t)] + \frac{\sum_{t=0}^{T-1}\sum_{j=t}^{T-1} \tau_j \|\mathbf{v}(\mathbf{x}_j)\|^2}{\lambda \underline{\eta}}.$$

Using Eq. (20) and $\eta_t \ge \underline{\eta} > 0$ for all $t \ge 1$, we have

$$\mathbb{E}[\|\mathbf{v}(\mathbf{x}_j)\|^2] \le \frac{\mathbb{E}[\|\mathbf{x}_j - \mathbf{x}^*\|^2] - \mathbb{E}[\|\mathbf{x}_{j+1} - \mathbf{x}^*\|^2]}{2(\lambda - \overline{\eta} - \tau\overline{\eta})\underline{\eta}}.$$

**Algorithm 2** Adaptive Online Gradient Descent with Noisy Feedback Information

**Initialization:** $\mathbf{x}_0 \in \mathbb{R}^n$, $\eta_1 = 1/\beta$ for some $\beta > 0$ and $r \in (1, +\infty)$.
**for** $t = 0, 1, 2, \ldots$ **do**
  **for** $i = 1, 2, \ldots, N$ **do**
    $x_i^{t+1} = x_i + \eta_{t+1}\mathbf{v}(\mathbf{x}_t)$.
    $\Delta\mathbf{x}_{t+1} = \sum_{j=0}^t \eta_{j+1}^{-2}\|\mathbf{x}_j - \mathbf{x}_{j+1}\|^2$
    $\eta_{t+2} = 1/\sqrt{\beta + \log(t+2) + \Delta\mathbf{x}_{t+1}}$.
  **end for**
**end for**

Putting these pieces with the fact that $\{\tau_t\}_{t\ge 0}$ is an nonincreasing sequence yields that

$$\begin{aligned}\sum_{t=0}^{T-1}\sum_{j=t}^{T-1} \tau_j \|\mathbf{v}(\mathbf{x}_j)\|^2 \;&\le\; \left(\sum_{t=0}^{T-1}\tau_t\right)\left(\sum_{t=0}^{T-1}\|\mathbf{v}(\mathbf{x}_t)\|^2\right) \\ &\le\; \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|^2}{2(\lambda - \overline{\eta} - \tau\overline{\eta})\underline{\eta}}\left(\sum_{t=0}^{T-1}\tau_t\right).\end{aligned}$$

Together with the fact that $\mathbb{E}[\sum_{t=0}^T \epsilon(\mathbf{x}_t)] = O(1)$ (cf. Theorem 4.4), we have

$$\mathbb{E}[\epsilon(\mathbf{x}_T)] \;\le\; \frac{\sum_{t=0}^T \mathbb{E}[\epsilon(\mathbf{x}_t)]}{T+1} + \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|^2}{T+1}\left(\sum_{t=0}^{T-1}\tau_t\right).$$

This completes the proof. □

### 4.3. Adaptive OGD Learning

We study the convergence property of Algorithm 2 under the noisy model (2) with relative random noise (3) satisfying that there exists $a(t) = o(1)$ such that

$$\frac{\log(T+1)}{T+1}\left(\sum_{t=0}^{T-1}\tau_t\right) \;=\; O(a(T)). \tag{14}$$

Note that Eq. (14) is slightly stronger than Eq. (13).

**Theorem 4.7** *Fix a $\lambda$-cocoercive game $\mathcal{G}$ with continuous action spaces $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathbb{R}^{n_i}, \{u_i\}_{i=1}^N)$ with an nonempty set of Nash equilibrium $\mathcal{X}^*$. Under the noisy model (2) with relative random noise (3) satisfying Eq. (14), the adaptive noisy OGD iterate $\mathbf{x}_t$ satisfies*

$$\mathbb{E}[\epsilon(\mathbf{x}_T)] \;=\; \begin{cases} O(a(T)) & \text{if } a(T) = \Omega(\log(T)/T), \\ O(\log(T)/T) & \text{otherwise.} \end{cases}$$

The proof technique is new and can be interpreted as a novel combination of that in Theorem 3.4 and 4.6. We refer the interested reader to the appendix for the details.

## Acknowledgements

We would like to thank the area chair and three anonymous referees for constructive suggestions that improve the paper. Zhengyuan Zhou was supported by the IBM Goldstine Fellowship. This work was supported in part by the Mathematical Data Science program of the Office of Naval Research under grant number N00014-18-1-2764.

## References

Arora, S., Hazan, E., and Kale, S. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

Bach, F. and Levy, K. Y. A universal algorithm for variational inequalities adaptive to smoothness and noise. In *COLT*, pp. 164–194, 2019.

Bloembergen, D., Tuyls, K., Hennes, D., and Kaisers, M. Evolutionary dynamics of multi-agent learning: a survey. *Journal of Artificial Intelligence Research*, 53:659–697, 2015.

Blum, A. On-line algorithms in machine learning. In *Online algorithms*, pp. 306–325. Springer, 1998.

Blum, A. and Mansour, Y. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*. Cambridge university press, 2006.

Daskalakis, C., Ilyas, A., Syrgkanis, V., and Zeng, H. Training GANs with optimism. In *ICLR*, 2018.

Facchinei, F. and Pang, J.-S. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Science & Business Media, 2007.

Hall, P. and Heyde, C. C. *Martingale Limit Theory and Its Application*. Academic Press, 2014.

Hazan, E. *Introduction to Online Convex Optimization*. Foundations and Trends(r) in Optimization Series. Now Publishers, 2016. ISBN 9781680831702. URL https://books.google.com/books?id=IFxLvgAACAAJ.

Hazan, E., Agarwal, A., and Kale, S. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, December 2007.

Kalai, A. and Vempala, S. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Krichene, S., Krichene, W., Dong, R., and Bayen, A. Convergence of heterogeneous distributed learning in stochastic routing games. In *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*, pp. 480–487. IEEE, 2015.

Lam, K., Krichene, W., and Bayen, A. On learning how players learn: estimation of learning dynamics in the routing game. In *Cyber-Physical Systems (ICCPS), 2016 ACM/IEEE 7th International Conference on*, pp. 1–10. IEEE, 2016.

Mertikopoulos, P. and Zhou, Z. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1-2):465–507, 2019.

Mertikopoulos, P., Belmega, E. V., Negrel, R., and Sanguinetti, L. Distributed stochastic optimization via matrix exponential learning. 65(9):2277–2290, May 2017.

Mertikopoulos, P., Papadimitriou, C., and Piliouras, G. Cycles in adversarial regularized learning. In *SODA*, pp. 2703–2717. SIAM, 2018a.

Mertikopoulos, P., Papadimitriou, C., and Piliouras, G. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2703–2717. SIAM, 2018b.

Monnot, B. and Piliouras, G. Limits and limitations of no-regret learning in games. *The Knowledge Engineering Review*, 32, 2017.

Palaiopanos, G., Panageas, I., and Piliouras, G. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30*, pp. 5872–5882. Curran Associates, Inc., 2017.

Polyak, B. T. *Introduction to Optimization*, volume 1. Optimization Software Inc., Publications Division, New York, 1987.

Quanrud, K. and Khashabi, D. Online learning with adversarial delays. In *Advances in Neural Information Processing Systems*, pp. 1270–1278, 2015.

Rosen, J. B. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pp. 520–534, 1965.

Shalev-Shwartz, S. and Singer, Y. Convex repeated games and Fenchel duality. In *Advances in Neural Information Processing Systems 19*, pp. 1265–1272. MIT Press, 2007.

Shalev-Shwartz, S. et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Shoham, Y. and Leyton-Brown, K. *Multiagent Systems: Algorithmic, Game-theoretic, and Logical Foundations*. Cambridge University Press, 2008.

Viossat, Y. and Zapechelnyuk, A. No-regret dynamics and fictitious play. *Journal of Economic Theory*, 148(2):825–842, 2013.

Zhou, Z., Mertikopoulos, P., Bambos, N., Glynn, P., and Tomlin, C. Multi-agent online learning with imperfect information.

Zhou, Z., Mertikopoulos, P., Bambos, N., Boyd, S., and Glynn, P. W. Stochastic mirror descent in variationally coherent optimization problems. In *Advances in Neural Information Processing Systems*, pp. 7040–7049, 2017a.

Zhou, Z., Mertikopoulos, P., Bambos, N., Glynn, P. W., and Tomlin, C. Countering feedback delays in multi-agent learning. In *NIPS '17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017b.

Zhou, Z., Mertikopoulos, P., Moustakas, A. L., Bambos, N., and Glynn, P. Mirror descent learning in continuous games. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 5776–5783. IEEE, 2017c.

Zhou, Z., Mertikopoulos, P., Athey, S., Bambos, N., Glynn, P. W., and Ye, Y. Learning in games with lossy feedback. In *Advances in Neural Information Processing Systems*, pp. 5134–5144, 2018.

Zhou, Z., Mertikopoulos, P., Bambos, N., Boyd, S. P., and Glynn, P. W. On the convergence of mirror descent beyond stochastic convex programming. *SIAM Journal on Optimization*, 30(1):687–716, 2020a.

Zhou, Z., Mertikopoulos, P., Moustakas, A., Bambos, N., and Glynn, P. Robust power management via learning and game design. *Operations Research*, 2020b.

Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *ICML '03: Proceedings of the 20th International Conference on Machine Learning*, pp. 928–936, 2003.